DOCUMENT RESUME

ED 369 653 SE 054 379

AUTHOR Altschuld, James W.; Kumar, David D.
TITLE Program Evaluation Models for Science Education: A Synthesis of Literature Sources.
PUB DATE Mar 94
NOTE 16p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (Anaheim, CA, March 26-29, 1994).
PUB TYPE Reports - Research/Technical (143) -- Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Research; Elementary Secondary Education; Evaluation Methods; *Evaluation Research; Literature Reviews; Models; *Program Evaluation; *Science Education; Science Programs; Student Evaluation

ABSTRACT

A literature search of the ERIC system was conducted with regard to papers focused on evaluation models in science education. The purposes were to identify models, their common and unique features, and to examine the guidance they could provide for evaluating science education programs. The search yielded a small number of entries pertinent to the areas of interest which were subsequently classified into three categories--micro (formative through the completion and summative evaluation of programs and products) and macro (contextual) levels and the interface between them. The categories were discussed along with some of the more general evaluation literature dealing with evaluation models. Conclusions were then drawn about why only a limited literature base existed, general models versus a more "situation" imbedded approach to understanding evaluation, the value of model type thinking, and implications for the future needs of evaluation in science education. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 369 653

Program Evaluation Models for Science Education: A Synthesis of Literature Sources

James W. Altschuld

The National Center for Science Teaching and Learning
The Ohio State University
1929 Kenny Road
Columbus, OH 43210

David D. Kumar

Florida Atlantic University
College of Education
2912 College Avenue
David, FL 33314

Summary of a Poster Presentation at the Annual Meeting of the National Association for Research in Science Teaching, Anaheim, CA, March 1994

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David D. Kumar

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE054379

Abstract

A literature search of the ERIC system was conducted with regard to papers focussed on evaluation models in science education. The purposes were to identify models, their common and unique features, and to examine the guidance they could provide for evaluating science education programs. The search yielded a small number of entries pertinent to the area of interest which were subsequently classified into three categories - micro (formative through the completion and summative evaluation of programs and products) and macro (contextual) levels and the interface between them. The categories were discussed along with some of the more general evaluation literature dealing with evaluation models. Conclusions were then drawn about why only a limited literature base existed, general models versus a more "situation" imbedded approach to understanding evaluation, the value of model type thinking, and implications for the future needs of evaluation in science education.

Introduction

There has been rising national concern for what is being taught in American schools and the quality of learning and instruction. Science education has been involved in the conceptualization and implementation of reform via such activities as the Project 2061 (Rutherford & Ahlgren, 1990), and the establishment of a national agenda for research in science education (Shymansky, 1992). A concomitant of this concern and activity is the need for systematic approaches or models for the evaluation of change and new directions in education and science education, in particular.

Seeking literature regarding the conceptualization and design of evaluations for science education programs and projects represents a complex undertaking. In this paper, one part of a larger, on-going search strategy is reported. Its specific focus is on models of evaluation for science education with, at times, reference to the general evaluation literature. To this effect the questions addressed in this study are: What models of program evaluation in science education are described in the literature?; What are their common features?; and What guidance do they provide for evaluating science education programs? The general evaluation literature contained many illustrations of models with the 1960's and early 1970's representing an especially productive period for this type of conceptual endeavor. The 1960's and early 1970's were characterized by both the attempt to reform education and the influx of money into the educational system through such legislation as the Elementary and Secondary Education Act of 1965 (ESEA), the Vocational Educational Amendments of 1963 and 1968, and the Education for All Americans Act of 1975. With increases in funding came an expanded cry for greater accountability and better ways in which to evaluate programs. The writings and evaluation perspectives of Cronbach (1963), Cronbach and others (1980), Scriven (1971), Stake (1967), and Stufflebeam (1971) became standard fare for courses in educational evaluation (see Worthen & Sanders, 1987; Worthen & Sanders, 1991; Altschuld & Thomas, 1991). Additional general models for program evaluation have been suggested in other fields such as training and development (Kirkpatrick, 1960; Brinkerhoff, 1987), the Extension Service (Bennett, 1972), the four A model in mental health (cited in Mathews, 1985) as well as newer approaches to program evaluation in education (Patton, 1990; Guba & Lincoln, 1989).

These conceptualizations exemplify what Borich (1983) referred to as **models of science** as opposed to **models in science**. A model of science is a general rationale or tool that helps to guide thinking and serves as a heuristic device in organizing thoughts. It is more amorphous and its form may vary somewhat depending on the vantage point that is taken. Models of science aid in thinking about general processes and attack strategies for resolving problems. They are facilitators of thoughts or as Borich termed them, "persuasions".

Using a model of science perspective, the literature review was limited to publications and materials that dealt with the heuristic dimensions or the generic processes used to design program evaluation in science education. (Sources that described the details of evaluating a specific project or program but that did not take the generic evaluation perspective were assigned to the second aspect of the search strategy mentioned previously.) By using this approach and the models so located, it might be possible to identify and highlight not solely models but the features of programs which would be of most benefit to evaluate. In the subsequent sections of this paper the search strategy, categories of sources that emerged, substantive aspects contained in them, and a discussion of themes and issues inherent in the reviewed literature will be provided.

Sampling: The domain of the literature search was program evaluation models in science education. To gather a body of knowledge reflecting such models, published studies and reports from 1966 to 1991 were targeted for the search. Sampling was based on a library title approach using the ERIC database. The descriptors 'Evaluation,' 'Assessment,' 'Model(s),' and 'Science Education' were utilized to examine the ERIC system for entries in accord with standard ERIC procedures (Houston, 1990). The term 'assessment', however, was dropped from the

descriptors because an initial review revealed an extensive use of it in relation to classroom testing and performance assessment studies, and thus it would tend to produce stray samples not within the scope of the investigation.

The search resulted in 34 studies ranging in content from models to various aspects of evaluation such as teacher evaluation. Therefore, further screening was carried out in two stages. Stage I: the 34 studies were analyzed independently by each of the two researchers for studies reporting program evaluation models in science education. The abstract of each sample was read and then classified into one of two categories -- concepts of "program evaluation models" and "other." Via this process, a total of 19 studies, six journal articles and 13 ERIC documents including dissertations were identified. Stage II: the sample of studies collected through Stage I was subjected to more indepth review of the content of the complete articles using the classification criteria mentioned earlier. This led to a final set of eight usable studies consisting of three journal articles (Welch, 1974; Mayer & Stoever, 1978; Espejo, Good & Westmeyer, 1975; Exline & Tonelson, 1987; Exline, 1985) and five ERIC documents (Cheu et al. 1979; Henkin et al., 1979; Pines, 1980; Shell et al., 1986; & Small, 1988). In addition, a secondary search from one of the journal articles (Exline & Tonelson, 1987), and a professional exchange with another science education researcher yielded two more (secondary) journal articles. The remaining 11 items from Stage I fell into Teacher Evaluation and Process Variables, which are not within the scope of this text.

Categories of Sources: Surprisingly, as noted in the search strategy only a small (n = 10) number of sources seemed pertinent to the nature of the investigation. Although other writers undoubtedly have dealt with and referred to evaluation concepts and models, the stringent

search criteria eliminated most of these sources in favor of those where evaluation modeling was the primary focus. When the located papers, presentations and reports were further scrutinized, several categories of evaluative ideas were apparent. First, was the concept of a micro-developmental or what might be thought of as the traditional formative view of evaluation that occurs during the production of science education curricula and/or projects. It placed major emphasis on the evaluation of the steps or stages inherent in the development of products and the achievement of outcomes specified for them. Second, there was another subset of papers that focussed on two interrelated themes, one was evaluation of the (macro-system level) context in which science education programs are imbedded and the other was, to some degree, the evaluative interface between the micro-developmental and the macro-system levels. In several of these writings stress was placed on the fact that science education programs were indeed highly dependent on certain contextual characteristics for successful implementation.

Evaluation: At the Micro-Developmental (Formative) Level: Evaluation, primarily of a formative nature regarding the systematic development of science education curricula, products, and processes, was most evident in the writings of Pines (1980), Mayer and Stoever (1978), and, to a lesser degree, Small (1988). A three phase developmental approach was noted in both Pines' (1980) Audio-Tutorial Elementary Science Project (A-TESP) and Mayer and Stoever's (1978) Crustal evaluation Project. Small (1988) reported on a two-phased evaluation of an instructional program.

In Pines, the **first phase** was pre-developmental with the objectives of deciding upon the content of A-TESP and determining its potential effectiveness to teach certain science concepts. Mayer and Stoever tried out lesson modules with students and subjected the content of lessons to

expert review by participating scientists. Although Small's (1988) ideas did not represent a comprehensive model for program evaluation in science education they do have bearing on this discussion. Across these authors, as in other more generalized evaluation models developed in the 1960's and 1970's, the initial emphasis was primarily in relationship to the process of forming value judgments about the product to be generated and ways in which possible outcomes could be assessed.

In the **second phase** of Pines, greater stress was placed on collecting data from the implementation of a **more finished** piece of curriculum. Pines used this phase to revise the content of the A-TESP, to ameliorate its deficiencies, to determine appropriate levels or standards of learning to be achieved by students, and to gain an understanding of the "gross limitations" of the product in the field. Almost identically, Mayer and Stoever examined and verified content, gathered student background information, and obtained teacher perceptions and opinions in conjunction with the actual instructional objectives that the teachers held for the educational activities. Ascertaining the quality of the objective tests developed in phase one, was another part of their evaluation. For Small the strategy at this point was instrument development.

Obviously, in **phase three**, the impact/effectiveness of the product (summative evaluation) would be the main consideration based upon the perspective that products had been conceptualized well and refined during previous periods. Therefore, Pines estimated the effect of A-TESP on concept acquisition through summative evaluation of participating children. Mayer and Stoever, expanded the testing of crustal evolution modules to a more representative sample of students along with the use of comparison groups in their third phase. All of the

above evaluation activities took place under conditions of rigorous experimental control with the classes of participating teachers being supervised closely by the local project staff. (Small, as noted, did not utilize a third phase.)

While the micro-developmental (formative) level of evaluation was prominent in Pines, Mayer and Stoever, and Small as highlighted in this review, it wasn't solely confined to those authors. Welch, to a limited extent, implied some of the same types of processes in his 1974 article. Virginia's Standards of Learning-Science (SOL-Science) by Exline (1985) and the Evaluation of a Child-Structured Curriculum by Espejo, Good and Westmeyer (1975) have some similar developmental components. The writing of

spejo, Good and Westmeyer also is related but had a different conceptual base for the evaluation of the child structured science curriculum employing theories of intellectual development to guide the eneration f evaluation instruments for monitoring and assessing the intellectual progress of children and clearly was more focussed at the micro level of evaluation.

Evaluation: At the Macro-System (Contextual) Level: The macro-system level of evaluative thought is prominent in the work of Exline (1985) and Exline and Tonelson (1987). It was somewhat emphasized or referred to in the writing of Welch (1977) and the ideas of Shell, Horn, and Severs (1986). At the macro-system level, Exline (1985) described the development of a comprehensive approach begun in 1978 in the State of Virginia to improve the effectiveness of science education programs. As suggested previously, 13 Standards of Learning (SOL) - Science. With the comprehensive scope of the SOL-Science, its accompanying materials, and the changes it represented for all of science education in Virginia, the State recognized that there would have to be a prolonged "buy in" period for administrators, teachers, parents and other

stakeholders. Thus, six critical support components (administrative support, community involvement, learning environment, teacher preparation, fixed facilities and instructional materials) were identified that must be in place for programs to be successful. In 1987, Exline and Tonelson, provided an overview of the subsequent development of what was called the Science Education Program Assessment Model (SEPAM) with in-depth approach to determining the degree to which the six support components were "addressing the thirteen standards" (p.72) that formed the basis of Virginia's reform movement in science education. One of its key premises is that quality in science education is a function of the level to which science education is supported and perceived positively by relevant constituencies.

In examining issues of measurement Welch (1974) noted that two foci should be considered -- measurement of individuals and measurement for overall program evaluation purposes. By separating these measurements and especially by describing the latter one, Welch was, in effect, directing attention toward the idea of the overall program perspective and that decisions regarding it were located within a larger, more complex context that would have impact on final decisions. Such an evaluation would not only be of value in understanding the system but also in determining needs. Context evaluation was a major part of his CIPP (Context, Input, Process, and Product) Model of evaluation.

Evaluation: The Interface Between the Micro-Developmental and Macro-System Levels: What seems to be neglected in evaluation model considerations and in the literature identified from the search strategy is an explicit examination of the exchange of information or the interface between the micro and macro levels. Shell and his co-authors argued that in macro or system evaluation, the question of the nature and integrity of the treatment is always a matter

of concern. With this perspective as a framework, the authors created the WE DO - THEY DO model of evaluation. It consisted of a series of procedures that would assist a program or project in determining responses to three questions: Did the staff do what they said they were going to do?, Did participants do what they said they were going to do? and If everyone fulfilled on what was expected of them in the first two questions how would we know if the program did any good or accomplished anything? By utilizing this model and detailed steps for each of the questions, expectations of actors would be clarified, the relationship of actions to outcomes would become explicit and direct as opposed to implicit and indirect, and the measurement of both process implementation and resultant outcomes would be significantly improved. Regarding the interface and linkage aspect, the authors presented an interesting example. They noted that in juvenile delinquency programs, recidivism is a frequently stated and evaluated outcome of delinquency programs at perhaps the higher system level, while "there is little direct relationship between program activities and recidivism" (p.12). The work on SEPAM and Welch's views dealt with it but the aspects of interface were more inferred by the authors of this manuscript than being singled out by those whose efforts were being reviewed. It is stressed here because it is assumed to be of relatively high importance for the development of both a good heuristic model of evaluation and guidelines for evaluating what is done in science education.

Conclusions and Discussion: The **first conclusion** is that there is a dearth of literature focussing specifically on evaluation models in science education. Either evaluation models are not of importance, past writings in evaluation modeling in science education continue to be sufficient (as judged by the dates of some of the references located in the course of this

investigation), general evaluation models in education provide enough guidance for evaluating science education programs, evaluation as viewed broadly from a general model perspective is perceived to have relatively limited payoff for science education at this time, or there is dissatisfaction with the general model approach. All are possible explanations for the lack of directly applicable citations. Another explanation regarding these results relates to the feeling or perception that general models will always be deficient given the situational nature of evaluation. Jackson (1990), in commenting about the "situation" or "context" imbedded nature of most research and evaluation, has suggested that despite this observation it is legitimate to seek general approaches or modes of thought. Moreover, other fields such as training and development, mental health, and extension view generic models tailored to their fields as useful for guiding the evaluation endeavor. They serve the heuristic function presented earlier.

A **second conclusion**, closely related to the first one is that the search strategy adapted for this review was flawed. This is based upon the assumption that there are model related writings embedded in the context of actual product and program evaluations. Key words in their abstracts may have focussed primarily upon actual evaluation activities rather than the concept of model. Therefore there could be more "models in use" or more discussions of model concepts than uncovered by our search strategy. (The second part of the literature review currently underway should prove useful in this regard.) Conversely and equally plausible is the possibility that **a sizeable portion** of the literature on evaluation models was located through the search strategy. If the examination or building of models was important it would have been more prominent in titles and abstracts and more emphasized by authors. The literature base, especially as related to the concept of models of science education evaluation, may not be there.

A **third** conclusion arises from the press for alternative, authentic, portfolio and similar assessment procedures that commands much of the current evaluative attention in virtually all areas of education including science education. While this focus is important and may even be part of a perceived dissatisfaction with general evaluation model thinking, it directs most of our vision toward outcomes and somewhat away from the broader perspective of an overall evaluation model that explains the systemic context of education. Narrowing the focus has advantages but the restriction of range may come with costs. Thus, while we appreciate the benefits of the new wave of assessment, we recommend not limiting the evaluation of science education programs and that the field should consider the fuller description and understanding of programs and projects that would be afforded from the reexamination and use of program evaluation models.

The **final conclusion,** in the absence of recent thinking about overall models, is that there is merit in revisiting older ones and, perhaps, in developing new general conceptualizations. The **model of science** is a valuable guide to thought and action.

Bibliography

Altschuld, J. W., & Thomas, P. M. (1991) The teaching of evaluation: 25 years of change and growth. [In J. W. Altschuld (Ed.) Educational evaluation: An evolving field]. Theory Into Practice, XXX(1), 22-29.

Bennett, C. F. (1972). A hierarchy of program evaluation applied to volunteer leadership development. Washington, D.C.: federal Extension Service. (ERIC Document Reproduction Service No. ED 073 370)

Borich, G. D. (1983). Evaluation models: A question of purpose not terminology. Educational Evaluation and Policy, 5(1), 61-63.

Brinkerhoff, R. O. (1987). Achieving results from training. San Francisco: Jossey-Bass.

Cronbach, L. J. (1963). Course improvement through evaluation. Teachers College Record, 64, 672-683.

Cronbach, L. J., Ambron, S.R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. Toward reform of program evaluation. San Francisco: Jossey-Bass.

Cheu, J., Iozzi, L., Maul, J. P., & Henkin, G. (1979). The socio-scientific reasoning model: Instruments for evaluation. A paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta, GA, March 1979.

Espejo, M., Good, R., & Westmeyer, P. (1975). Evaluation of a child-structured science curriculum using the intellectual models of Piaget and Guilford. Journal of Research in Science Teaching, 12(2), 147-155.

Exline, J. D. (1985). Virginia's long range plan for science education. NSTA's Supplement of Science Education Suppliers.

Exline, J. D., & Tonelson, S. W. (1987). Virginia's science education program assessment model resource guide. NSTA's Supplement of Science Education Suppliers.

Guba, E. G., & Lincoln, Y.S. (1989). Effective evaluation. San Francisco: Jossey-Bass.

Henkin, G., Iozzi, L., Cheu, J., & Maul, J. (1979). An evaluation of curriculum materials based upon the socio-scientific reasoning model. A paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta, GA, March 1979.

Houston, J. E., (Ed.). (1990). Thesaurus of ERIC descriptors, 12th edition. Phoenix, AZ: Oryx Press.

Jackson, P.W. (1990). The functions of educational research. Educational Researcher, 19(1), 3-9.

Kirkpatrick, D. L. (1960). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 14(1), 13-18.

Mathews, W. D. (1985). Assessing the four A's. In the Evaluation SourceBook of the Ohio Program Evaluators' Group, pp. 3-20. Columbus: Ohio Department of Mental Health.

Mayer, V. J., & Stoever, Jr., E. C. (1978). NAGT crustal evolution education project: A unique model for science curriculum materials development and evaluation. Science Education, 62(2), 173-179.

Patton, M. Q. (1990). Qualitative evaluation and research methods. Beverly Hills: Sage Publications.

Pines, L. A. (1980). A model for program development and evaluation: The formative role of summative evaluation and research in science education. A paper presented at the annual conference of the International Congress for Individualized Instruction, Windsor, Canada, October 1980.

Rutherford, F.J., & Ahlgren, A. (1990). Science for All Americans. New York: Oxford University Press.

Scriven, M. (1971). The methodology of evaluation. In D. L. Stufflebeam (Ed.), Educational evaluation: Theory and practice, (pp. 60-106). Belmont, California: Wadsworth Publishing Company, Inc.

Shell, D. F., Horn, C. A., & Severs, M. K. (1986). We do--they do: A model for practical service program evaluation. A paper presented at the annual meeting of the American Evaluation Association, Kansas City, MO, October 1986.

Shymansky, J. A. (1992). Establishing a research agenda: Critical issues of science curriculum reform. Journal of Research in Science Teaching, 29(8), 749-748.

Small, L. (1988). Science process evaluation model. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 1988.

Stake, R. E. (1967). The countenance of educational evaluation. Teachers College Record, 68, 523-540.

Stufflebeam, D. L. (1968, January). Evaluation as enlightenment for decision-making. Paper presented at a working conference on assessment theory sponsored by the Association for Supervision and Curriculum Development, Sarasota, FL.

Stufflebeam, D. L. (1971). Excerpts from evaluation as enlightenment for decision-making. In D. L.

Stufflebeam (Ed.), Educational evaluation: Theory and practice, (pp.143-150). Belmont, California: Wadsworth Publishing Company, Inc.

Welch, W. W. (1974). The process of evaluation. Journal of Research in Science Teaching, 11(3), 175-184.

Worthen, B. R., & Sanders, J. R. (1987). Educational evaluation: Alternative approaches and practical guidelines. New York: Longman.

Worthen, B. R., & Sanders, J. R. (1991). The changing face of educational evaluation. [In J. W. Altschuld (Ed.), Educational evaluation: An evolving field]. Theory Into Practice, XXX(1), 3-12.